Air Quality Analysis

Problem Statement

The project aims to analyze and visualize air quality data from monitoring stations in Tamil Nadu. The objective is gain insights into air pollution trends, identify areas with high pollution levels, and develop a predictive model to estimate RSPM/PM10 levels based on 502 and NO2 levels. This project involves defining objectives, designing the analysis approach, selecting visualization techniques, and creating a predictive model using Python and relevant libraries.

Abstract

Air pollution is a major environmental problem that has significant impacts on human health and the environment. Air quality monitoring and prediction are essential for developing and implementing effective air pollution mitigation strategies.Machine learning (ML) is a powerful tool that can be used to analyze and predict complex data. ML models have been used to predict a variety of environmental phenomena, including air quality.This project aims to develop an ML-based model to analyze and predict air quality. The model will be trained on a historical dataset of air quality data, including pollutant concentrations, meteorological data, and other relevant factors. The trained model will be used to predict air quality for future time periods and to identify the factors that most influence air quality.

Objectives

Collect and prepare a historical dataset of air quality data.
Develop and train an ML model to analyze and predict air quality.
Evaluate the performance of the trained model on a held-out test set.
Identify the factors that most influence air quality.
Develop a user-friendly interface for the model to make it accessible to decision-makers and the public.
The findings from this project will be used to develop and implement effective air pollution mitigation strategies. The project will also contribute to the development of ML-based air quality models that can be used in other cities and regions around the world.

Analysis Approach

- Collect and prepare the data. This will involve gathering a historical dataset of air quality data, including pollutant concentrations, meteorological data, and other relevant factors.
- The data will need to be cleaned and preprocessed to make it suitable for machine learning.
- Choose a machine learning model. There are a variety of machine learning models that can be used for air quality prediction. Some popular choices include linear regression, support vector machines, and neural networks. The choice of model will depend on the specific characteristics of the data and the desired accuracy of the predictions.
- Train the model. Once the model has been chosen, it needs to be trained on the historical dataset. This will involve feeding the data to the model and allowing it to learn the relationships between the different variables.
- Evaluate the model. Once the model has been trained, it needs to be evaluated on a held-out test set. This will give an indication of how well the model will generalize to new data.
- Deploy the model. Once the model has been evaluated and found to be satisfactory, it can be deployed to production. This may involve developing a user-friendly interface for the model or integrating it into an existing air quality monitoring system.

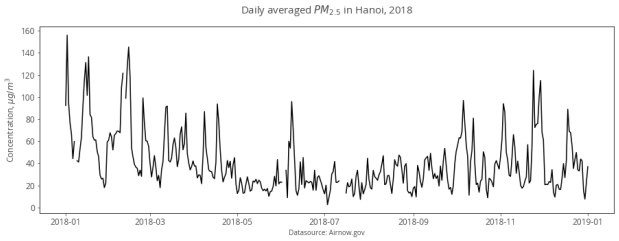
In addition to the above steps, it is also important to consider the following:

Feature selection. Not all of the features in the dataset will be equally important for predicting air quality. It is important to select the most relevant features to improve the accuracy of the model.
Model hyperparameter tuning. The hyperparameters of a machine learning model control its behavior. It is important to tune the hyperparameters to optimize the performance of the model.
Model interpretation. It is important to understand why the model is making the predictions that it is. This can be done by interpreting the model's parameters and weights.
Model monitoring. It is important to monitor the performance of the deployed model over time. This is because the factors that influence air quality can change over time.
By following these steps, you can develop a robust and reliable machine learning-based model for air quality analysis and prediction.

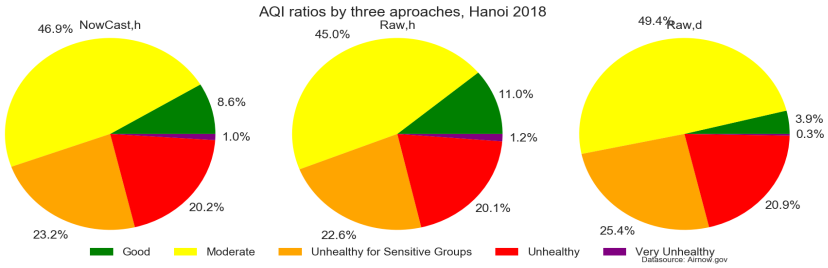
Visualization Selection

The selection of visualizations for an air quality analysis and prediction project will depend on the specific goals of the project and the target audience. Here are some examples of visualizations that could be used for an air quality analysis and prediction project:

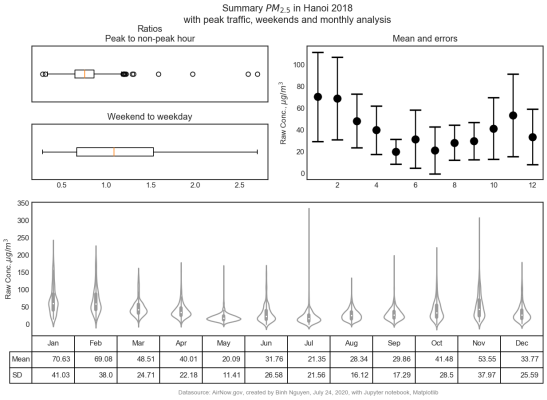
Line charts: Line charts can be used to show how air quality changes over time.



Pie chart with Air Quality Index (AQI)

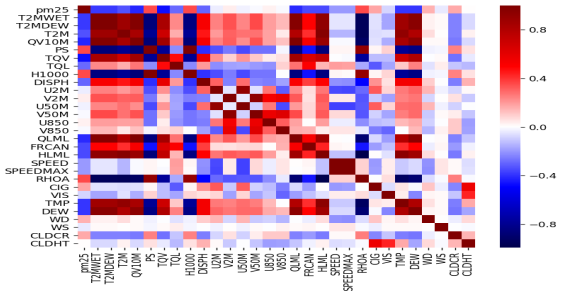


Correlation of PM2.5

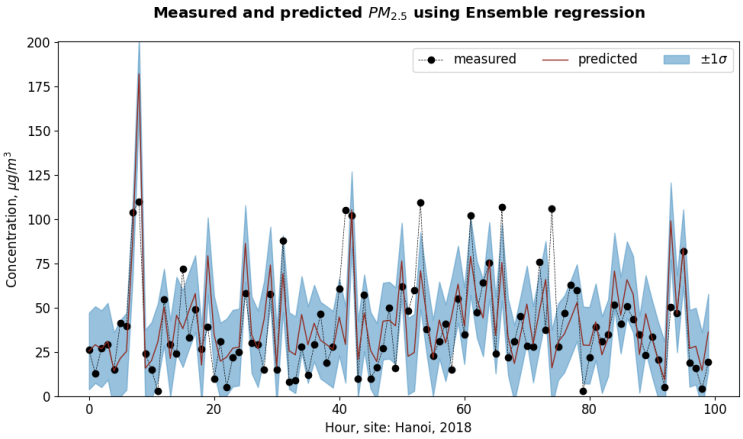


Data selection:

Correlation Map :



Ensemble Regression :



Tools & Packages :

Python (3.6.9)
Matplotlib (3.1.2)
pandas (1.1.0)
Seaborn
MetPy
scikit-learn
scipy